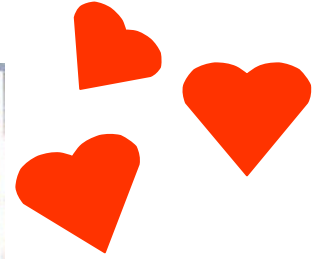


# Pathema Stuff

Feb 6<sup>th</sup>, 2006

# Anthrax SNPs




- Sequenced 7 *B. anthracis* strains
    - Including reference, Ames Ancestor, complete
  - Load into Chado
  - Employ mummer pipeline
  - Identify polymorphic sites
  - Curation at level of electropherogram
- Total of 3,523 unique SNPs discovered

# Anthrax SNPs

List SNPs - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://pathema.tigr.org/tigr-scripts/pathema/sybil/shared/list\_snps\_ Go

 **Pathema**  
Bioinformatics Resource Center

Search Annotation for Go

Home | Genomes | Searches | Comparative Analyses | Gene Lists | Carts | Downloads | Links

## List SNPs: find all SNPs for a specific reference sequence and target strain(s)

**Example SNP reports:**

- ♦ All pairwise SNPs between *B. anthracis str. Kruger B* (assembly 1046) and *B. anthracis Ames Ancestor* (assembly 2479)
- ♦ A single polymorphic base on *B. anthracis str. Kruger B* assembly 1046, position 1910
- ♦ Number of SNPs per gene for *B. anthracis Ames Ancestor* assembly 2479

**1. Select a single reference sequence:**

gb21\_76\_assembly - 5.23 Mb (Bacillus anthracis A0039) ▼

**2. Check one or more target organism(s)/strain(s) to include:**

☐ *Bacillus anthracis str. France*  
☐ *Bacillus anthracis Vollum*  
☐ *Bacillus anthracis str. A01055*  
☐ *Bacillus anthracis Western North America USA6153*  
☐ *Bacillus anthracis A0039*  
☐ *Bacillus anthracis Ames Ancestor*  
☐ *Bacillus anthracis str. Kruger B*

All SNP alignments between the reference sequence and the selected organism(s)/strain(s) will be displayed. If no target strain(s) are chosen then all SNPs on the reference sequence will be displayed.

**3. Choose a report type:**

List all pairwise SNPs between reference and query ▼

**4. Select an output format:**

HTML table ▼

**5. Run the report:**

List SNPs or Reset the form

Questions? Comments? [Contact Us](#)  
Pathema Version 1.4 January 13, 2006  
©1999-2006 The Institute for Genomic Research

# Anthrax SNPs

list SNPs - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://pathema.tigr.org/tigr-scripts/pathema/sybil/shared/list\_snps.cgi?site=pathe

Search Annotation for Go

Pathema  
Bioinformatics Resource Center

Home | Genomes | Searches | Comparative Analyses | Gene Lists | Carts | Downloads | Links

**Pairwise SNP report: 57 SNP alignments found for reference seq=gak\_1046\_assembly , target seq=gb6\_2479\_assembly**

Reference				Target				Synonymous or Nonsynonymous
sequence	position	type	AA info	sequence	position	type	AA info	
gak_1046_assembly	148 G	intergenic		gb6_2479_assembly	81614 a	intergenic		
gak_1046_assembly	286 C	intergenic		gb6_2479_assembly	81752 t	intergenic		
gak_1046_assembly	1910 G	intergenic		gb6_2479_assembly	83376 a	coding	2479_GBAA_pX02_0092 :: R (1)	
gak_1046_assembly	8175 T	intergenic		gb6_2479_assembly	89648 g	coding	2479_GBAA_pX02_0099 :: D (1)	
gak_1046_assembly	8261 G	intergenic		gb6_2479_assembly	89728 a	coding	2479_GBAA_pX02_0100 :: S (1)	
gak_1046_assembly	10302 A	intergenic		gb6_2479_assembly	91769 g	coding	2479_GBAA_pX02_0102 :: A (3)	
gak_1046_assembly	10386 T	intergenic		gb6_2479_assembly	91853 c	coding	2479_GBAA_pX02_0102 :: V (3)	
gak_1046_assembly	11409 G	intergenic		gb6_2479_assembly	92876 a	coding	2479_GBAA_pX02_0102 :: S (3)	
gak_1046_assembly	17466 C	intergenic		gb6_2479_assembly	4105 g	coding	2479_GBAA_pX02_0007 :: Y (3)	
gak_1046_assembly	21387 C	intergenic		gb6_2479_assembly	8103 t	coding	2479_GBAA_pX02_0012 :: A (3)	
gak_1046_assembly	22163 T	intergenic		gb6_2479_assembly	8942 c	coding	2479_GBAA_pX02_0013 :: T (3)	
gak_1046_assembly	22174 G	intergenic		gb6_2479_assembly	8953 a	coding	2479_GBAA_pX02_0013 :: L (1)	
gak_1046_assembly	22178 G	intergenic		gb6_2479_assembly	8957 a	coding	2479_GBAA_pX02_0013 :: N (3)	
gak_1046_assembly	22186 A	intergenic		gb6_2479_assembly	8965 t	coding	2479_GBAA_pX02_0013 :: I (1)	
gak_1046_assembly	22190 G	intergenic		gb6_2479_assembly	8969 a	coding	2479_GBAA_pX02_0013 :: I (3)	
gak_1046_assembly	22198 G	intergenic		gb6_2479_assembly	8977 t	coding	2479_GBAA_pX02_0013 :: R (1)	
gak_1046_assembly	22219 T	intergenic		gb6_2479_assembly	8998 c	coding	2479_GBAA_pX02_0013 :: G (1)	
gak_1046_assembly	22223 C	intergenic		gb6_2479_assembly	9002 t	coding	2479_GBAA_pX02_0013 :: V (3)	
gak_1046_assembly	22229 A	intergenic		gb6_2479_assembly	9008 g	coding	2479_GBAA_pX02_0013 :: G (3)	
gak_1046_assembly	22232 C	intergenic		gb6_2479_assembly	9011 t	coding	2479_GBAA_pX02_0013 :: Q (3)	
gak_1046_assembly	22235 T	intergenic		gb6_2479_assembly	9014 c	coding	2479_GBAA_pX02_0013 :: Q (3)	
gak_1046_assembly	22244 T	intergenic		gb6_2479_assembly	9023 a	coding	2479_GBAA_pX02_0013 :: T (3)	
gak_1046_assembly	22245 T	intergenic		gb6_2479_assembly	9024 g	coding	2479_GBAA_pX02_0013 :: T (2)	
gak_1046_assembly	22246 C	intergenic		gb6_2479_assembly	9025 t	coding	2479_GBAA_pX02_0013 :: T (1)	
gak_1046_assembly	22247 T	intergenic		gb6_2479_assembly	9026 a	coding	2479_GBAA_pX02_0013 :: S (3)	

# Anthrax SNPs

list SNPs per gene - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://pathema.tigr.org/tigr-scripts/pathema/sybil/shared/list\_snps.cgi?site=pathema&ref\_s=

Search Annotation for Go

Home | Genomes | Searches | Comparative Analyses | Gene Lists | Carts | Downloads | Links

**List SNPs: 40 genes with SNPs found on reference seq=gb6\_2479\_assembly**

organism	sequence	gene/ORF	product	SNPs	S	N	U	O	S/N
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0013</a>	membrane protein, putative, (pXO2-14)	22	0	0	0	22	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0045</a>	surface-layer N-acetylmuramoyl-L-alanine amidase, (pXO2-42)	6	0	0	0	6	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0060</a>	gamma-glutamyltranspeptidase, (pXO2-55)	5	0	0	0	5	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0092</a>	hypothetical protein, (pXO2-73)	3	0	0	0	3	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0102</a>	hypothetical protein, (pXO2-81)	3	0	0	0	3	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0039</a>	plasmid replication protein, (pXO2-38)	3	0	0	0	3	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0066</a>	transcriptional repressor PagR, (pXO1-109)	2	0	0	0	2	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0101</a>	hypothetical protein, (pXO2-80)	2	0	0	0	2	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0057</a>	capsule synthesis trans-acting positive regulator, putative, (pXO2-53)	2	0	0	0	2	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0061</a>	capsule biosynthesis protein CapA, (pXO2-56)	2	0	0	0	2	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0028</a>	hypothetical protein, (pXO2-29/30)	2	0	0	0	2	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0032</a>	hypothetical protein, (pXO2-33)	2	0	0	0	2	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0100</a>	conserved hypothetical protein, (pXO2-79)	2	0	0	0	2	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0007</a>	NLP/P60 family protein, (pXO2-08)	2	0	0	0	2	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0044</a>	hypothetical protein	1	0	0	0	1	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0062</a>	capsule biosynthesis protein CapC, (pXO2-57)	1	0	0	0	1	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0035</a>	conserved hypothetical protein, (pXO2-35)	1	0	0	0	1	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0064</a>	hypothetical protein	1	0	0	0	1	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0005</a>	hypothetical protein, (pXO2-06)	1	0	0	0	1	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0016</a>	conserved domain protein, (pXO2-17)	1	0	0	0	1	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0077</a>	capsule synthesis trans-acting positive regulator, (pXO2-64)	1	0	0	0	1	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0024</a>	hypothetical protein, (pXO2-26)	1	0	0	0	1	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0029</a>	hypothetical protein, (pXO2-31)	1	0	0	0	1	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0068</a>	signal peptidase I, (pXO2-59)	1	0	0	0	1	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0033</a>	hypothetical protein, (pXO2-34)	1	0	0	0	1	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0091</a>	hypothetical protein	1	0	0	0	1	
<i>B. anthracis</i> Ames Ancestor	gb6_2479_assembly	<a href="#">2479_GBAA_pXO2_0089</a>	FunZ family protein, (pXO2-71)	1	0	0	0	1	

# Anthrax SNPs

[illegible]

# Anthrax SNPs: Near future

- Affymetrix resequencing chip
- 1,000 Anthrax strains
- Implications:
  - Performance ?
  - Data model explosion?
  - Virtual sequence

# Orthology Displays

protein cluster 14064660 - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://pathema.tigr.org/tigr-scripts/pathema/sybil/shared/show\_protein\_cluster.cgi?site=pat Go

**Pathema**  
Bioinformatics Resource Center

Search Annotation for Go

Home | Genomes | Searches | Comparative Analyses | Gene Lists | Carts | Downloads | Links

## Protein cluster 14064660

### CLUSTER SUMMARY

**Algorithm:** cogs  
**Description:**  
**Number of proteins/genes:** 5  
**Average BLASTP identity:** 99.0%  
**Average BLASTP coverage:** 96.7%

### CLUSTERED PROTEINS/GENES

1.	[ <i>B. anthracis Vollum</i> ]	<a href="#">gb20_84_ORF05339</a>	BAVO5339	228 aa	1 exon(s)	sodium/alanine symporter family protein
2.	[ <i>B. anthracis str. Kruger B</i> ]	<a href="#">gak_1043_ORF05340</a>	BAKB5340	201 aa	1 exon(s)	sodium/alanine symporter family protein
3.	[ <i>B. anthracis Western North America USA6153</i> ]	<a href="#">gb4_2017_ORF05202</a>	BAWN5202	201 aa	1 exon(s)	sodium/alanine symporter family protein
4.	[ <i>B. anthracis str. A01055</i> ]	<a href="#">gb7_273_ORF05308</a>	BAGC5308	201 aa	1 exon(s)	sodium/alanine symporter family protein
5.	[ <i>B. anthracis A0039</i> ]	<a href="#">gb21_76_ORF05190</a>	BAAU5190	215 aa	1 exon(s)	sodium/alanine symporter family protein

### GENOMIC CONTEXT OF CLUSTERED GENES/ORFS

display at most: **5 kb** 10 kb 15 kb 25 kb 50 kb on either side of the highlighted genes:

gb20\_84\_assembly/- (5214542 bp)

4.97M

gb20\_84\_ORF05344 gb20\_84\_ORF05341 gb20\_84\_ORF05337 gb20\_84\_ORF05335

gb20\_84\_ORF05343 gb20\_84\_ORF05342 gb20\_84\_ORF05340 gb20\_84\_ORF05336 gb20\_84\_ORF05334

gb20\_84\_ORF05339 gb20\_84\_ORF05338

gak\_1043\_assembly/- (5197923 bp)

4.92M

gak\_1043\_ORF05346 gak\_1043\_ORF05343 gak\_1043\_ORF05339 gak\_1043\_ORF05337

gak\_1043\_ORF05345 gak\_1043\_ORF05342 gak\_1043\_ORF05338 gak\_1043\_ORF05336

gak\_1043\_ORF05344 gak\_1043\_ORF05341

gb4\_2017\_assembly/- (5236046 bp)

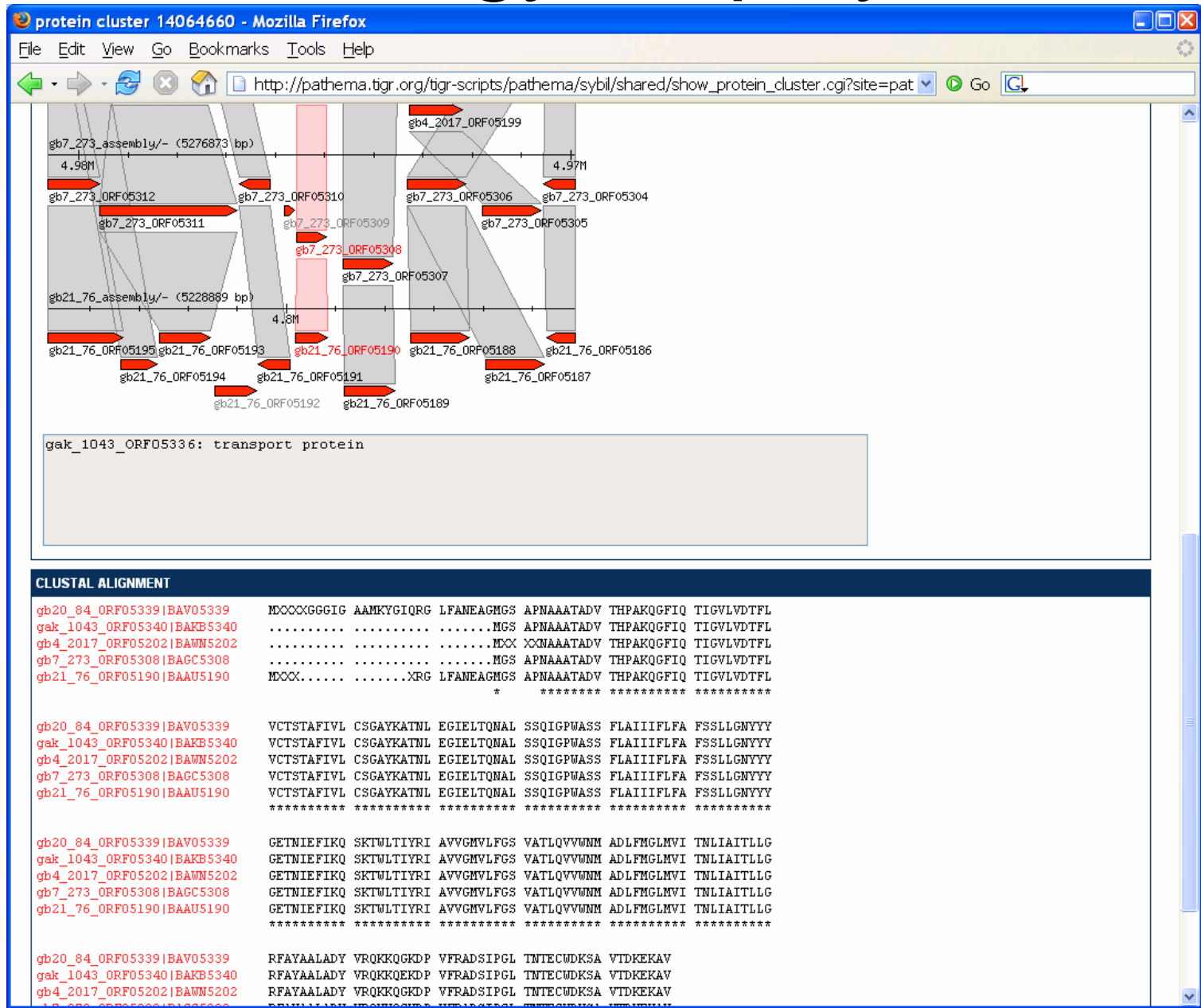
4.82M

gb4\_2017\_ORF05207 gb4\_2017\_ORF05204 gb4\_2017\_ORF05201 gb4\_2017\_ORF05198

gak\_1048\_ORF05340



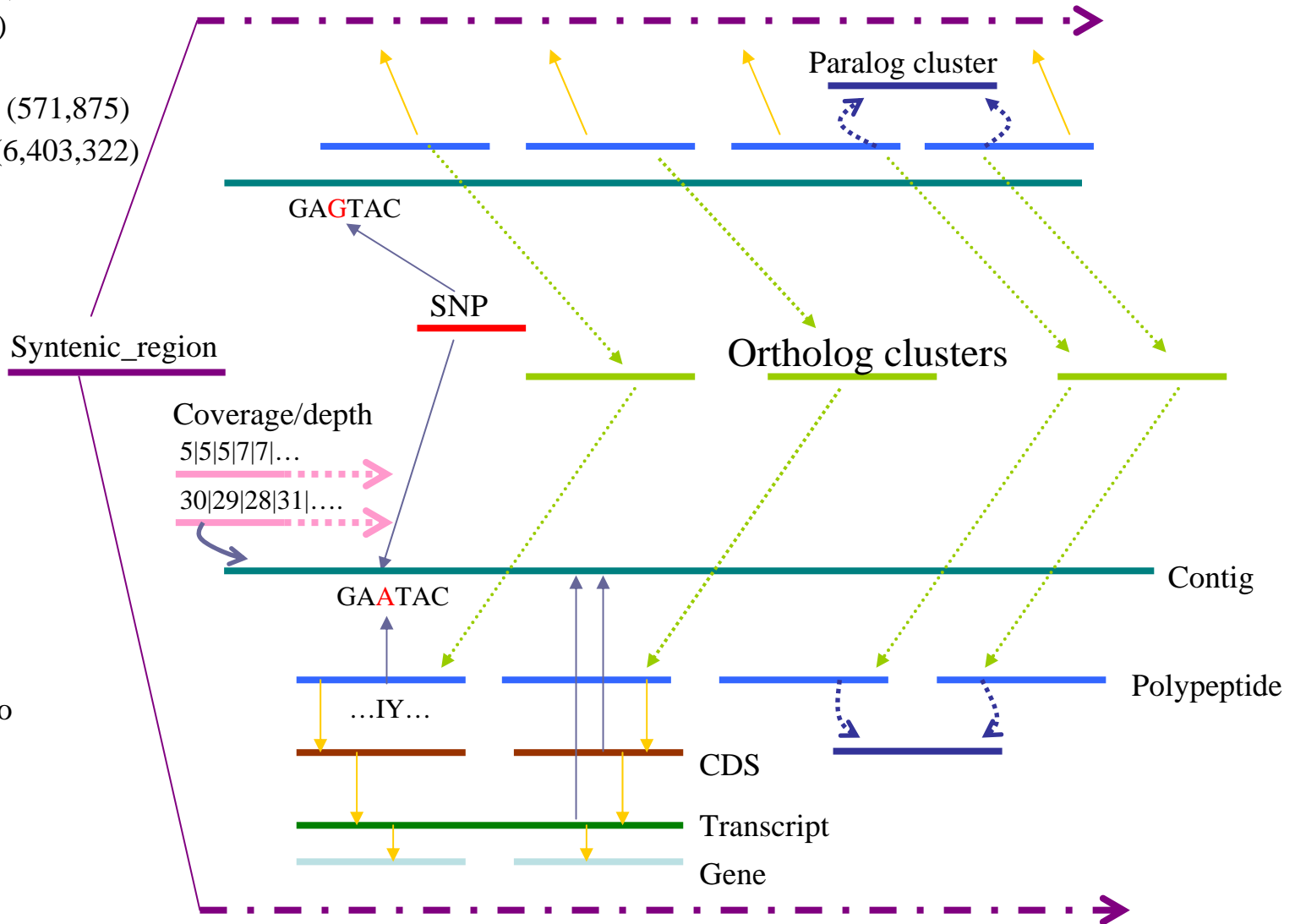
# Orthology Displays



## Features

- contigs (354)
- genes (149247)
- orthologs (5992)
- paralogs (1794)
- SNPs (4036)
- HMM matches (571,875)
- Blast matches (6,403,322)

# Chado features



# Reactions

Species	Genes	EC #s	Reactions	Pathways
B. Anthracis	1,619	656	943	75
B. Mallei	1,582	536	842	74
B. Pseudomallei	1,185	471	749	76
C. Botulinum	874	303	687	72
C. Perfringens	939	289	681	64



About

TOC

User Guide

Data Model

Schema

Extended search

Pathfinder

Download

Linking

Citing

## Metabolic Pathway, Reactions and Related Genes

Pathways	Loci Names	EC numbers	Gene Names	
1- and 2-Methylnaphthalene degradation	CPE0097	1.3.99.-	acdS acyl-CoA dehydrogenase	Naphthyl-2-methyl-succ
1- and 2-Methylnaphthalene degradation	CPE0449	1.1.1.1	adh alcohol dehydrogenase	(2-Naphthyl)methanol +
1- and 2-Methylnaphthalene degradation	CPE0449	1.1.1.1	adh alcohol dehydrogenase	1-Hydroxymethylnaphth
Alanine and aspartate metabolism	CPE1668	4.3.2.2	purB adenylsuccinate lyase	N6-(1,2-Dicarboxyethyl)-
Alanine and aspartate metabolism	CPE1668	4.3.2.2	purB adenylsuccinate lyase	N6-(1,2-Dicarboxyethyl)-
Aminoacyl-tRNA biosynthesis	CPE0014	6.1.1.11	serS serine-tRNA ligase	ATP + L-Serine + tRNA(
Aminoacyl-tRNA biosynthesis	CPE0014	6.1.1.11	serS serine-tRNA ligase	ATP + L-Serine + tRNA(
Aminoacyl-tRNA biosynthesis	CPE2464	6.1.1.14	glyS glycine-tRNA ligase	ATP + Glycine + tRNA(G
Aminoacyl-tRNA biosynthesis	CPE2464	6.1.1.14	glyS glycine-tRNA ligase	ATP + Glycine + tRNA(G
Aminophosphonate metabolism	CPE0071	2.7.8.-	clsA probable cardiolipin synthase	CMP-2-aminoethylphos
Aminophosphonate metabolism	CPE0071	2.7.8.-	clsA probable cardiolipin synthase	CMP-2-aminoethylphos
Aminophosphonate metabolism	CPE0365	2.7.8.-	clsB cardiolipin synthase	CMP-2-aminoethylphos
Aminophosphonate metabolism	CPE0365	2.7.8.-	clsB cardiolipin synthase	CMP-2-aminoethylphos
Aminophosphonate metabolism	CPE0711	2.6.1.-	aspC aspartate aminotransferase	3-Phosphonopyruvate +
Aminophosphonate metabolism	CPE0754	2.7.8.-	clsC cardiolipin synthase	CMP-2-aminoethylphos
Aminophosphonate metabolism	CPE0754	2.7.8.-	clsC cardiolipin synthase	CMP-2-aminoethylphos
Aminophosphonate metabolism	CPE0854	2.6.1.-	aspC aspartate transaminase	3-Phosphonopyruvate +
Aminosugars metabolism	CPE0075		PTS	UDP-N-acetyluramate
Aminosugars metabolism	CPE0076		PTS	UDP-N-acetyluramate
Aminosugars metabolism	CPE0077		PTS	UDP-N-acetyluramate
Aminosugars metabolism	CPE0154		probable	UDP-N-acetyluramate
Aminosugars metabolism	CPE0184	5.1.3.9	nanP putative N-acylglucosamine-6-phosphate 2-epimerase	N-Acetyl-D-glucosamine
Aminosugars metabolism	CPE0185	4.1.3.3	nanA acylneuraminate lyase	N-Acetylneuraminate <=
Aminosugars metabolism	CPE0266		probable	UDP-N-acetyluramate
Aminosugars metabolism	CPE0353	1.1.1.158	murB UDP-N-acetylenolpyruvoylglucosamine reductase	UDP-N-acetyluramate
Aminosugars metabolism	CPE0981		probable	UDP-N-acetyluramate
Aminosugars metabolism	CPE1364		beta-N-acetylhexosaminidase	UDP-N-acetyluramate
Aminosugars metabolism	CPE1992		probable	UDP-N-acetyluramate
Aminosugars metabolism	CPE2131	3.1.4.-	rng ribonuclease G	N-Acetyl-D-mannosamii
Aminosugars metabolism	CPE2176	3.5.1.25	nagA N-acetylglucosamine-6-phosphate deacetylase	N-Acetyl-D-glucosamine
Aminosugars metabolism	CPE2196		UDP-N-acetylglucosamine	UDP-N-acetyluramate
Aminosugars metabolism	CPE2329		phosphoglucosamine	UDP-N-acetyluramate
Aminosugars metabolism	CPE2434	3.5.99.6	nagB glucosamine-6-phosphate isomerase	D-Glucosamine 6-phos
Aminosugars metabolism	CPE2490		glucosamine-1-phosphate	UDP-N-acetyluramate
Arginine and proline metabolism	CPE0150		2-dehydro-3-deoxyphosphogluconate/4-hydroxy-2-ox	L-Proline + NAD+ <=> (
Arginine and proline metabolism	CPE0150		2-dehydro-3-deoxyphosphogluconate/4-hydroxy-2-ox	trans-4-Hydroxy-L-prolin
Arginine and proline metabolism	CPE0256		probable	L-Proline + NAD+ <=> (
Arginine and proline metabolism	CPE0256		probable	trans-4-Hydroxy-L-prolin
Arginine and proline metabolism	CPE0551	3.5.3.11	speB probable agmatinase	Agmatine + H2O <=> P
Arginine and proline metabolism	CPE0757		creatinase	L-Proline + NAD+ <=> (
Arginine and proline metabolism	CPE0757		creatinase	L-Proline + NAD+ <=> (
Arginine and proline metabolism	CPE0757		creatinase	trans-4-Hydroxy-L-prolin
Arginine and proline metabolism	CPE1050		5'-methylthioadenosine/S-adenosylhomocysteine	L-Proline + NAD+ <=> (
Arginine and proline metabolism	CPE1050		5'-methylthioadenosine/S-adenosylhomocysteine	trans-4-Hydroxy-L-prolin
Arginine and proline metabolism	CPE1670	2.6.1.1	aspC aspartate transaminase	L-erythro-4-Hydroxygluts

# Genome Properties

Project	Biochemical Pathways	Protein Transport	Small Molecule Transport	Non-transport
<i>B. anthracis</i> Ames	53 (5)	3 (0)	6 (2)	47 (10)
<i>B. mallei</i>	65 (11)	6 (2)	7 (2)	58 (9)
<i>B. pseudomallei</i>	66 (11)	7 (3)	8 (1)	66 (14)
<i>C. botulinum</i>	35 (10)	4 (0)	5 (1)	46 (12)
<i>C. perfringens</i>	35 (13)	2 (1)	4 (2)	45 (9)

Properties with partial evidence are counted as positive assertions and denoted in parentheses.

Assessment of 85 biochemical pathways and 19 specific transport networks

Genome Properties: a system for the investigation of prokaryotic content.  
Bioinformatics. 2005 Feb 1;21(3):293-306.

# Genome Properties


Genome Properties Definition - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

← → ↺ × 🏠

http://pathema.tigr.org/tigr-scripts/pathema/shared/GenomePropDefinitior

Go

 **Pathema**  
Bioinformatics Resource Center

Search Annotation for  for

Home | Genomes | Searches | Comparative Analyses | Gene Lists | Carts | Downloads | Links

Summary

Related Properties

Pathema Results

## Genome Property Definition Page ?

Accession: [GenProp0014](#) | Name: polyketide biosynthesis, type I

Download

Accession	GenProp0014
Name	polyketide biosynthesis, type I
Description	<p>Allowed states: YES NO Polyketides are an important class of natural product which are assembled by the sequential addition of units containing at minimum one acetate group. Polyketides include many antibiotics and other compounds of biological and economic importance. Polyketides may be synthesized by large multi-domain proteins (Type I polyketide synthetases) which carry out a so-called thio-template program in which successive modules of domains add single units to the growing end of the polyketide chain, modify the newly added piece (usually a series of reductions) and pass the chain on to the next module. The overall mechanism is similar to that observed in the biosynthesis of fatty acids in animals via the Type I fatty acid synthases. By contrast, Type II polyketide synthetases, which are similar to the Type II fatty acid synthases found in bacteria and plants, consist of separate proteins for each of the various functions involved in the overall process. Plant chalcones and stilbenes are synthesized by yet a third style of synthetase in which a single condensing enzyme is active for all condensation steps. Many of the constituent domains of type I polyketide synthetases have been defined by HMM models. These include acyl transferase (PF00698), beta-ketoacyl-synthase (PF00109, PF02801), acyl carrier protein (PF00550) and thioesterase (PF00975) domains. Other commonly found domains which have not yet been represented by HMMs include ketoreductase, dehydratase and enoyl reductase domains. This genome property will be set to YES if any single gene in the genome contains domains corresponding to PF00698, PF00109, PF02801 and PF00550. (These correspond to the minimum set of domains for production of a polyketide chain.) Failing that condition, a YES will be set if two consecutive genes in combination contain all of these domains (as type I polyketide synthetases are sometimes found as multiple genes broken up between some, but not all, of the domains). Otherwise, the property will be set to NO. A NULL value indicates that the genome has not yet been evaluated with regard to this property.</p>
TIGR Role	Biosynthesis of natural products
Gene Ontology Term	<a href="#">GO:0030639</a> : polyketide biosynthesis (biological_process)

Components

Download

Name	Step Num	Required	Evidence
acyl trans, keto synth and thioest domains in the same ORF	AT+KS+TE	YES	<a href="#">PF00698 + PF00109 + PF02801</a> + (RULES)

# OBO Edit Usage

- Properties have parent-child relationship
- Components are children of property
- Evidence is child of component
- Genes found by evidence are children of evidence
- Load those relationships into OBO edit...
- Exchange format for GP?

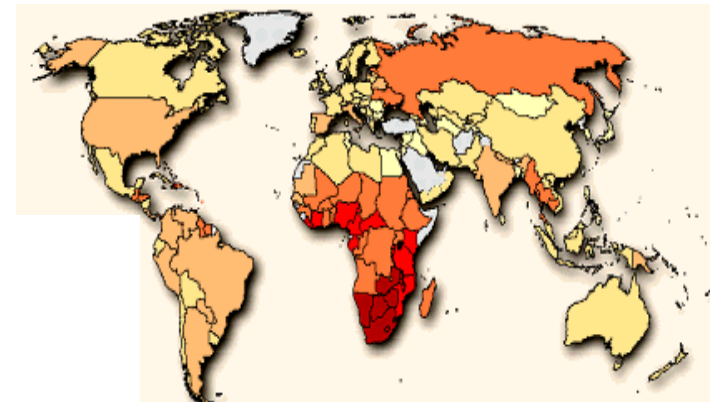






# Epidemiological data

- With more and more strains being sequenced we need to be able to correlate new data-types with sequence information....eg ...
  - Geographical information (country, city or GPS)
  - Clinical symptoms
  - Date of isolation
  - Host range (organism taxon, sex, age, etc)
  - Tissue/cell type that is infected
  - Method of transmission
  - reservoir
  - etc, etc ,etc



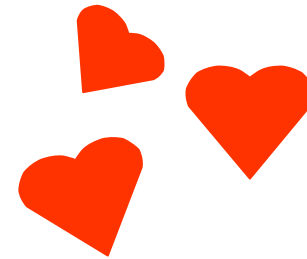
# Microbial Rosetta Stone Database

The Microbial Rosetta Stone Database: a compilation of global and emerging infectious microorganisms and bioterrorist threat agents.

Ecker DJ, Sampath R, Willett P, Wyatt JR, Samant V, Massire C, Hall TA, Hari K, McNeil JA, Buchen-Osmond C, Budowle B.

BMC Microbiol. 2005 Apr 25;5(1):19.

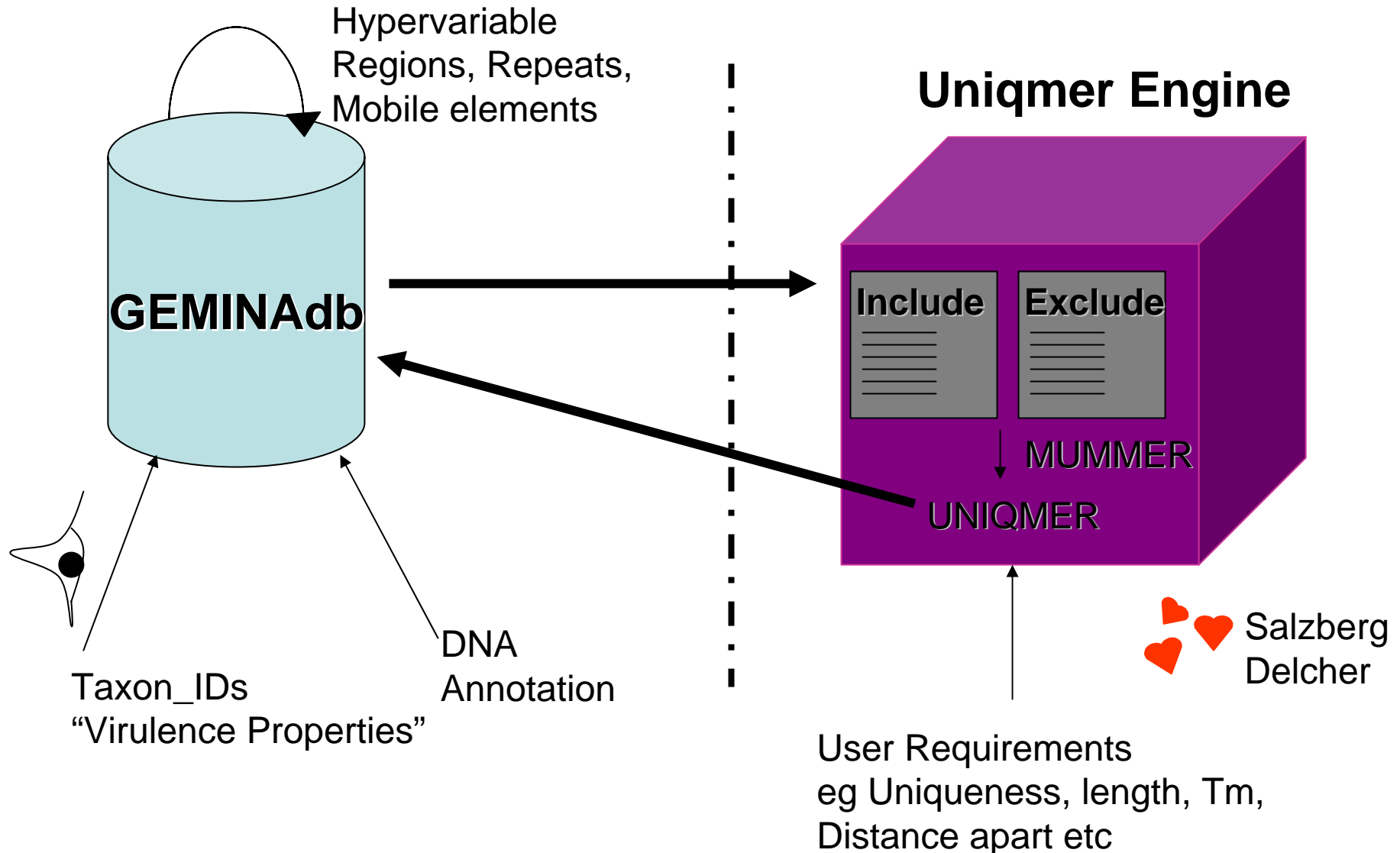
Ibis Therapeutics



# The Gemina database

- Fully relational
- Highly research-brittle
- IP agreements in progress
- Content
  - Protocols
  - Taxon→epi data
  - Hierarchically structured CVs:
    - anatomy, disease, symptom, transmission method, reservoir, geographic location
- Example queries:
  - “all of the DNA sequences of pathogens that infect the gut”
  - “all of the the airborne pathogens found in the USA”
  - “All the flu viruses isolated in 2005”

# The DNA signature pipeline

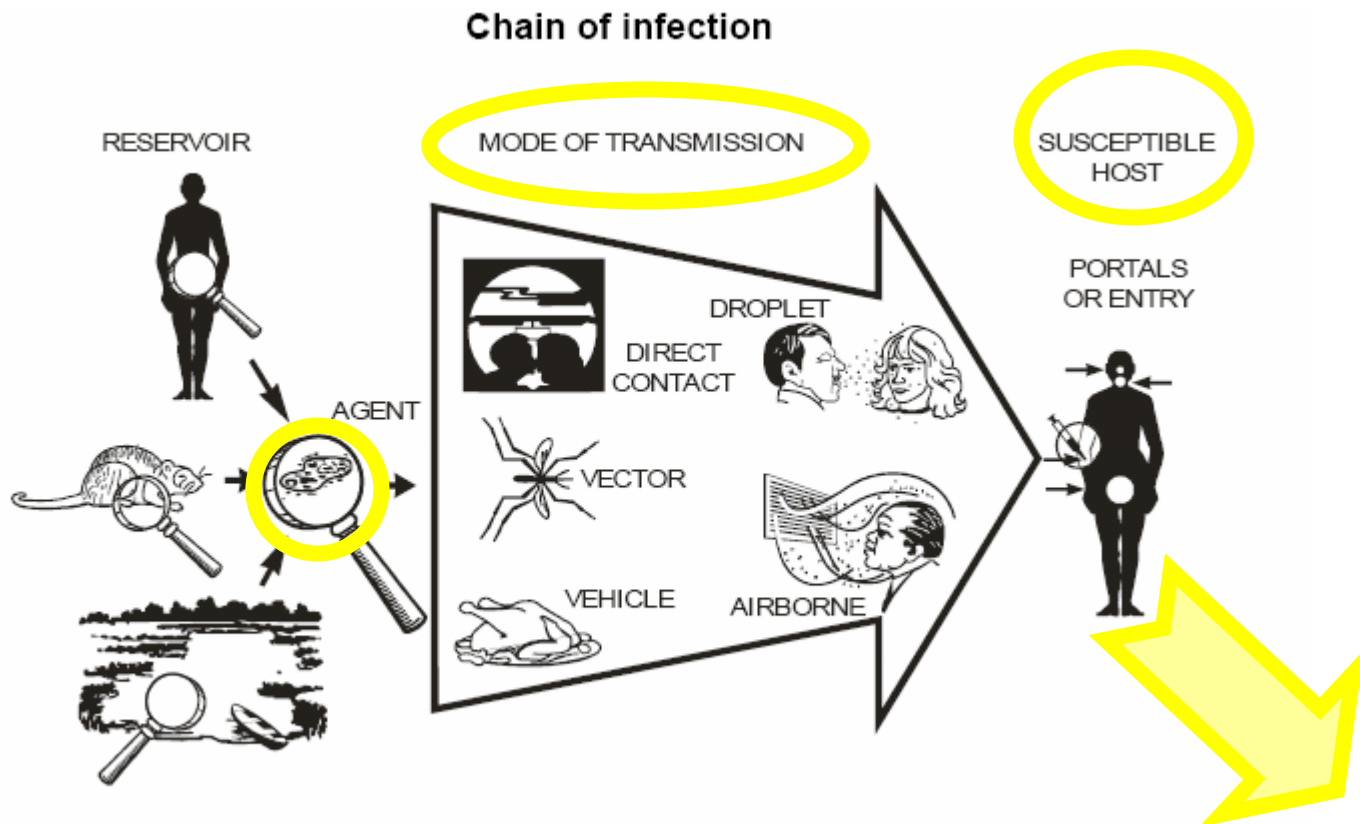


# Infection Systems : representation of the “Chain of Infection”

Pathogen -- Component -- Host -- Disease

*Bacillus anthracis str. Ames* – Skin – Human - Anthrax

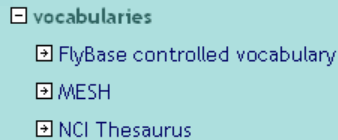
*Bacillus anthracis str. Ames Ancestor* – Respiratory Tract – Cow - Anthrax



# Ontologies & Controlled Vocabularies In Gemina

- infectious disease and body system oriented
- hierarchical query and retrieval

anatomy — disease — symptom — transmission method — reservoir — geographic location



## disease

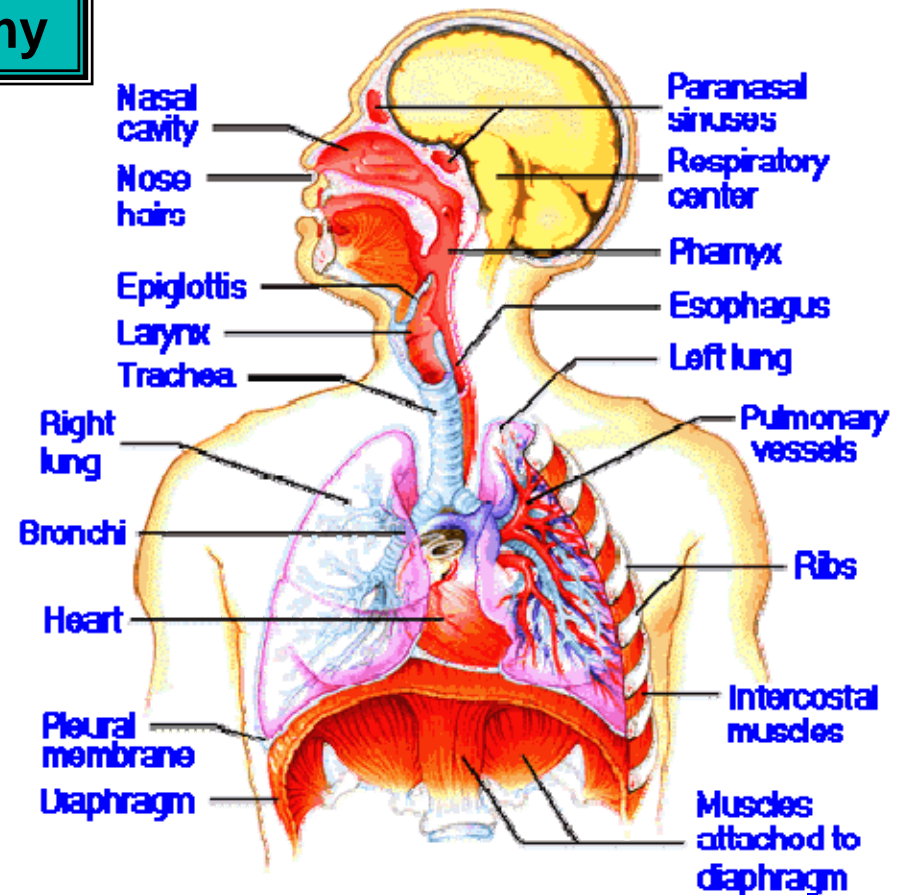
- +diseases of the respiratory system
- +infectious and parasitic diseases
  - +arthropod-borne viral disease
  - +intestinal infectious diseases
    - +other bacterial diseases
      - +bacterial infection
        - +gas gangrene
        - +staphylococcus infection
        - +tetanus

## reservoir

- +animal reservoir
  - +arthropod
  - +mollusc
- +environmental reservoir
  - +soil
  - +food
- +human reservoir
  - +blood
  - +respiratory tract

# Components -- Tissues -- Anatomy

- +Animal\_structure
- +Body\_region
- +Cardiovascular\_system
- +Cell
- +Digestive\_system
- +Embryonic\_structure
- +Endocrine\_system
- +Fluids\_and\_secretions
- +Hemic\_and\_immune\_system
- +Integumentary\_system
- +Musculoskeletal\_system
- +Nervous\_system
- +Respiratory\_system
- +Sense\_organ
- +Stomatognathic\_system
- +Tissue
- +Urogenital\_system



## Respiratory\_system

- + larynx
- + lung
- + pharynx
  - + nasopharynx
  - + oropharynx

# epidemiology CVs sources

- Mesh files at OBO:
  - anatomy (tissue component)
  - symptom
  - geographical location
- Rat Genome Database
  - disease → divided this into separate ontologies
    - diseases and symptoms.
- literature, web sites and epidemiology sources
  - transmission method
  - reservoir ontologies





# anatomy ontology (1,318)

- pared down ontology to describe
  - body parts
  - body systems
  - body components
- augmenting with specific human cell types.

# disease ontology (402)

- augmented
  - infectious disease
  - mycoses
  - zoonoses
- removed non-infectious conditions/diseases (e.g., congenital conditions).
- separated out symptom into separate ontology
- augmented this ontology with ICD-10 data the WHO



# symptom ontology (254)

- began at signs and symptoms node
- organized the symptoms ontology to match disease ontology
- separated:
  - General symptoms, v.
  - Symptoms involving the various body regions or systems

# geographical location (370)

- extracted from MeSH
  - defined the children of regions
  - includes additional nodes where needed (e.g., city/county)
  - added US Census
  - added CIA list of countries

# transmission method (11)

- curated transmission mode from Principles of Epidemiology, from CDC
  - direct modes
    - Kissing, skin-to-skin, contact w/ soil
  - indirect modes
    - Airborne
    - Vehicle borne, medical (tissue transplant)
    - Vector borne (flea bite)



# natural reservoir (81)

- created top nodes:
  - Environmental
  - Human
  - Animal
- added
  - literature
  - web sites
  - epidemiology sources.

# infection systems

anatomy — disease — symptom — transmission method — reservoir — geographic location

- Species (10)
  - Bacillus anthracis, Bacillus anthracis Ames, Bacillus anthracis Ames Ancestor, Bacillus anthracis Sterne (11)
  - Brucella spp, B. melitensis, B. abortus, B. suis, B. canis (85)
  - Burkholderia mallei: (29)
- Current total : 135
- Underway: Human and avian flu



Lynn Schriml

# Obligate Collaborations

- Anthrax SNPs: Jaques Ravel@tigr
- Chado: FlyBase and BDGP.
- Mummer/SNP/Signature: Salzberg@UMd
- Reactome: CSHL, EBI, GO<sup>tm</sup>
- Gene function CV: GO<sup>tm</sup> consortium
- MRS database: Ibis Therapeutics